Flexible, Eco-Friendly and Profitable Car Sharing Contracts: An Investigation into consumers' behavior

Paola Martin[1]

Abstract

Recently car sharing companies have grown across the United States. Members of these companies, when faced upon the decision on the amount of time to rent a car for, do not always make rational decisions. This behavior negatively affects both the customers and the company as if customers reserve a car for the wrong amount of time, they might either return it early and pay for extra time that they do not need, or they might return it late and incur a penalty fee. When this happens, the company needs to find a way to satisfy other customers' that might have been displaced. This research paper therefore aims at analyzing customers' behaviors and understanding which variables affect customers' decisions when reserving a car. Through a random effects logistic regression analysis, data was analyzed to understand which variables influence a late return. Results showed that the two variables playing the most critical roles in the percent change of late returns are the time of the day and whether or not a customer had to pay a penalty on their previous rental.

[1] I am grateful for the opportunity to do research under the supervision of Professor Diwakar Gupta.

1 Introduction

In recent years, car sharing has become a popular alternative to owning a car. Customers see this system as advantageous because it allows them to not having to look for a parking spot and not having to pay maintenance costs, while reducing pollution and congestion in the cities. This research analyzes a car sharing company[2] in the Midwestern region of the United States. This company offers a variety of contracts that vary in membership fees and per unit rental costs. Customers pay a membership fee in order to become members of the company and can then utilize its service. The booking system is fully online, where customers are able to see the availability of the cars and make reservations upon logging in to their accounts. Once the time of their reservation comes, they can just go to the desired location and use an activated fob to open the car. Members are also free to cancel their reservations, as long as that happens at least eight hours before the reserved start time. Additionally, members pay a penalty if they return the car more than five minutes late and may encounter fees for prohibited actions such as smoking in the car, causing a dead battery, not refilling the car when needed, and so on.

When deciding on the amount of time to reserve a car for, customers face a newsvendor decision problem. If they reserve the car for too long, they will have to pay for extra time that they do not need. In contrast, if they reserve the car for too short a time, they have to pay a penalty upon returning the car late. Much research has been done on the newsvendor model and its applications in operations management. In particular, research on the newsvendor model has focused on the mismatch cost parameters and on minimizing the expected cost. Another stream

---

[2] Due to privacy agreements the name of the company will not be mentioned.

of literature utilizes econometrics to obtain the cost parameters, under the assumption that decision makers act rationally and choose the optimal decision for a cost function that is unobservable to the researcher. This project utilizes a similar approach and offers an application of this model to a car sharing company. The objective of this project is to understand how customers make booking decisions and how particular customers' attributes affect their likelihood to return a car on time. In order to do so, a random effects regression model was used. This model was chosen to produce unbiased estimates of the coefficients and produce the smallest standard errors. Moreover, the effect of the variables is random and not constant, as in a fixed effect model. In this project, explanatory variables that describe customers' characteristics were identified in order to predict the likelihood of a particular customer to return the car late, as well as the probability of them to return more than fifteen minutes late. The result of this study will be utilized in the future to design rental contracts that lead to better social outcomes.

2 The Model

The newsvendor model is widely used in operations management to optimize inventory levels. In the model, the demand D is unknown and follows a distribution $F(\cdot)$. The decision maker selects a level q before demand is realized. If $q>D$, then the decision maker incurs the cost of overage, which is C_o. On the other side, if $q<D$, then the decision maker incurs a cost of underage, C_u. The objective is then to select a q that minimizes the expected cost, which is given by

$$\min_{q} \mathrm{E}\{Co\ (q - D) + \ + Cu\ (D - q) +\}. \tag{1}$$

The optimal solution satisfies the following condition

$$F(q^*) = \frac{C_u}{C_u + C_o}. \tag{2}$$

The right hand side of (2) is known as the critical ratio of the newsvendor model. If the decision maker were rational, they would choose the optimal q satisfying the above condition.

In order to understand what factors influence customers' decisions, a random effects logistic regression can be used. Given a vector X of covariates, which are variables affecting the relationship between dependent and independent variables, the probability of the dependent binary variable y to be non-zero is given by the following equation

$$P(y \neq 0|X) = G(\beta X + \varepsilon). \tag{3}$$

In equation (3), β is a vector of parameters, ε are i.i.d. Normally distributed error terms with mean 0 and standard deviation σ^2 and $G(t) = (1 - \exp(-t))^{-1}$. This model is used to understand how one unit change in an explanatory variable can affect the probability of y being different than zero. Our research focuses on the utilization of this structural model to understand the factors involved in the change from an on-time return to a late one.

3 Application to Car Sharing Company

This research aims at understanding consumers' behavior and the rationale behind their decisions making in the setting of a car-sharing company. Transactional data was obtained from the company for the purpose of this research for a period of 18 months. To start off, customers are divided into groups based on their age and location. It is believed that customers within a group, by virtue of having age and location similarities, act in a very similar way and face the same demand distributions. For each renter group *i,* customers incur a cost $C_{r,i}$, which is the per unit time rental cost, and $C_{p,i}$, which is the penalty cost. Therefore, if customers were risk-neutral and rational, they would all reserve the same amount of time according to the critical ratio

$$F(q^*) = \frac{C_{p,i}}{C_{p,i} + C_{r,i}}. \tag{4}$$

For each rental j in group i, let y_{ij} be a binary variable taking on value 0 if the customer returned the car on time and 1 otherwise. Then, according to the random effects logistic regression equation,

$$P(y_{ij} \neq 0 | X_{ij}) = G(\beta_0 + \beta_1 X_{ij}^1 + \beta_2 X_{ij}^2 + \cdots + \beta_k X_{ij}^k + \varepsilon_i). \quad (5)$$

In equation (5), X_{ij}^m represent attributes pertaining to customer j in group i and with m=1,…,k. When customers are rational, we should find that the hypothesis $\beta_1 = \beta_2 = \cdots = \beta_k = 0$ cannot be rejected.

Similarly, it is possible to find the probability of a customer returning the car more than fifteen minutes late, given the same explanatory variables X_{ij}. Let z_{ij} be a binary variable taking on value 1 if the customer returned the car more than fifteen minutes late and 0 otherwise. Then the probability of this late return is given by the equation

$$P(z_{ij} \neq 0 | X_{ij}) = G(\beta_0 + \beta_1 X_{ij}^1 + \beta_2 X_{ij}^2 + \cdots + \beta_k X_{ij}^k + \varepsilon_i). \quad (6)$$

Similarly to equation (5), in equation (6), X_{ij}^m represent attributes pertaining to customer j in group i and with m=1,…,k. Once again, when customers are rational, the hypothesis $\beta_1 = \beta_2 = \cdots = \beta_k = 0$ cannot be rejected. Equation (6) is helpful in predicting whether the late return is still within an acceptable time range, which is believed to be within fifteen minutes of the reservation end time.

The following section identifies the explanatory variables used in the research and how they were defined, given the data available. In addition, y and z are defined once again, along with the explanation of how those variables were calculated.

5 Description of Explanatory Variables

The explanatory variables used in the research can be grouped based on four categories. The first one consists of attributes related to the specific contracts offered by the company, the second one of customer's personal traits, the third one of customer's choices in renting a car and the fourth one of outside factors. In order to simplify the explanation, they are presented according to these four categories. The total number of observations available from the data is 10,874, while the total number of individual customers is 1,491. Out of these 10,874 rentals, 950 were cancelled and therefore these data was not further analyzed for the purpose of this research. The new number of observations is then 9,924 and the total number of customers is 1,403. The customer that reserved the most has a total of 127 rentals over the 18 months period. The graph below shows how many times different number of rentals occurred over the given period.

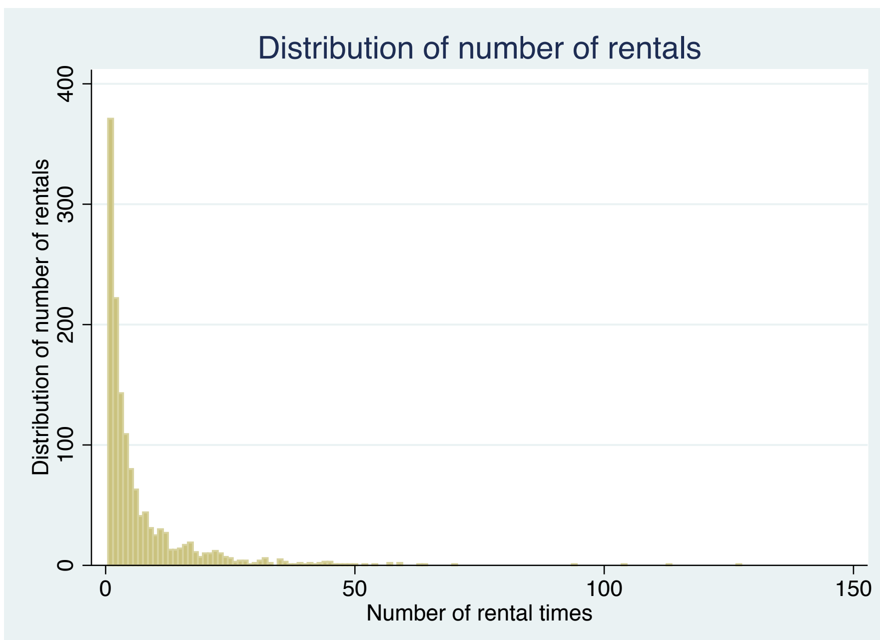

Figure 1: Distribution of number of rentals

The range of customers' rental goes from 1 to 127 times. In particular, according to the graph, 371 customers have a history of one rental, 222 of two and 143 of three in the 18 months period that the data was given.

Variables related to the specific contracts offered by the company are the type of plan and the per unit rental time. The company offers three main contracts to its customers, namely plans “A”, “B” and “C”. Plan “A” is available to a pre-selected group of customers that have an affinity with academia. This plan does not require a monthly fee and has an $8/hour rental cost. Customers are free to choose between plans “B” and “C”. The first one requires a $10/month membership fee and a $12/hour rental cost. Finally, plan “C” requires a $20/month membership fee and a $9/hour rental cost.

- Type of plan (*mymembershipid*)

The variable describing the type of contracts a customer can choose is categorical and assumes values 0 under plan “A”, 10 under “B” and 20 under “C”. Additionally, the given data included a few observations under some historical plans that have been grandfathered into some customers’ contracts. Because these plans together cover roughly 3% of the total number of observations available and they are not available to potential members at the time of doing this analysis, they are not taken into account in this analysis and will therefore not be considered in the remaining of the paper. The total number of observations now is then 9,633 and the total number of customers is 1,347. The table below indicates the number of observations for the three plans considered, along with the number of customers per each plan and the percent of late returns within each plan. It also shows on average, how many times customers rent a car for within each plan.

	Number of customers	Percent of total customers	Number of observations	Percent of total observations	Percent of late return within the plan	Average number of rentals per plan
Plan A	561	41.65%	4,434	46.03%	9.88%	8
Plan B	525	38.98%	2,905	30.16%	7.50%	5-6
Plan C	261	19.37%	2,294	23.81%	6.54%	8-9

Table 1: Type of plan

Table (1) shows that "Plan A" has the most number of observations, as well as the most number of customers. Additionally, customers under plan "A" tend to reserve an average of 8 times, those under plan "B" five to six times and those under plan "C" eight to nine times over the same amount of time. Also, the percent of late returns under plan "A" is higher than in the other two plans.

- Per unit rental time (C_r)

This variable was given values based on the rental cost of each plan. It therefore was assigned an 8 if under Plan "A", 12 if under "B" and 9 if under "C". The same number of observations and customers in this category can be found on Table (1).

The second group of variables pertains to particular characteristics of customers. These include age, location, state that issued the driving license and how long a customer has been a member of the company.

- Age (*myageyears*)

This numerical variable takes into account the age of the members. It is found by creating a variable containing a date on which this analysis was performed, in this case March 6th, 2015. The age is then found by subtracting the customer's birthdate from this date. The range of ages in the data goes from 19 years old to 82. The histogram below shows the frequency of each age year.

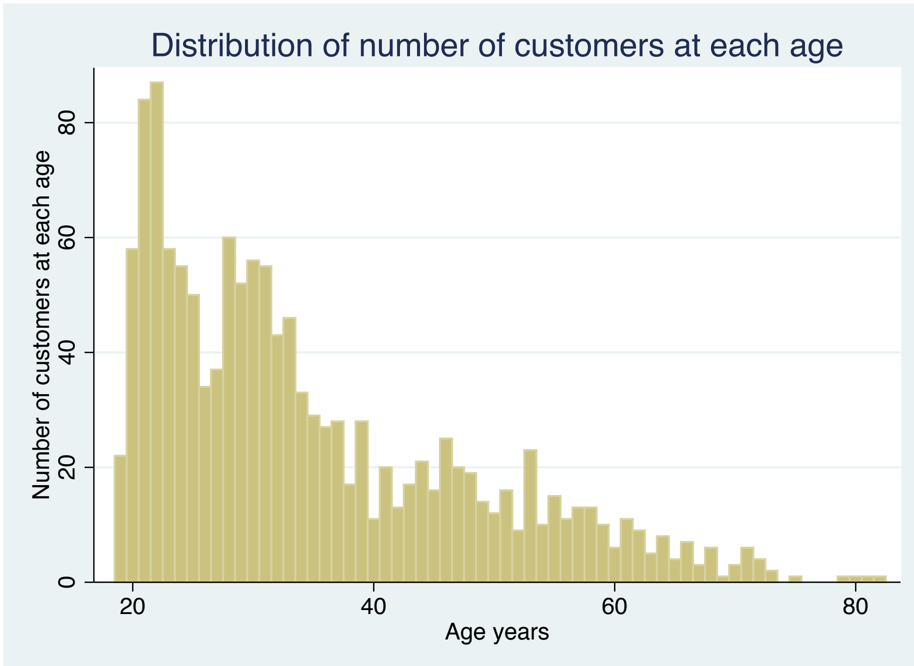

Figure 2: Distribution of number of customers at each age

From Figure (2), 71.20% of the customers are between 19 and 39 years old. The graphs below show the distribution of age for each plan.

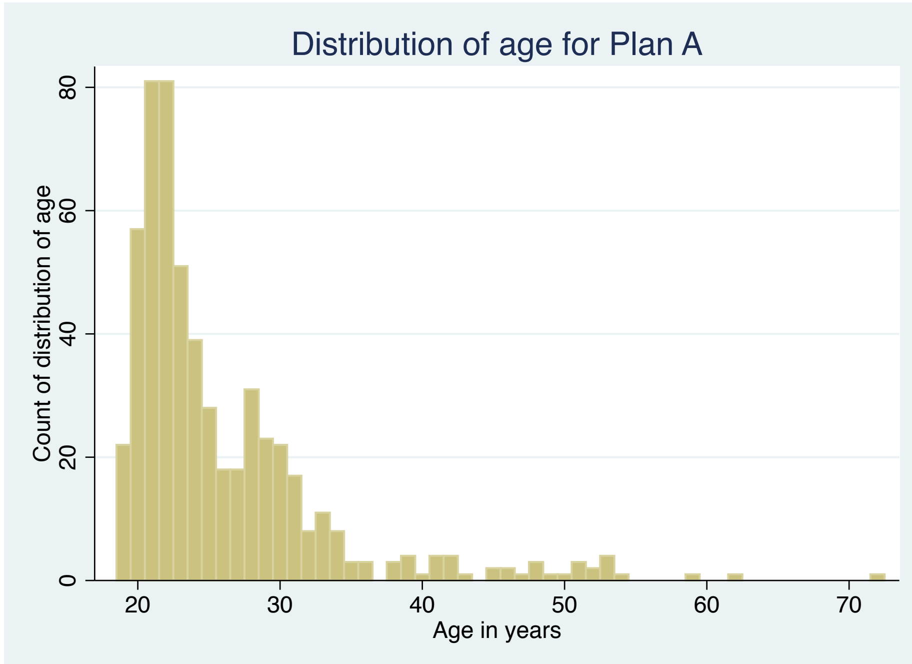

Figure 3: Distribution of age for Plan A

The majority (52.05%) of customers choosing plan “A” range between 19 and 23 years old.

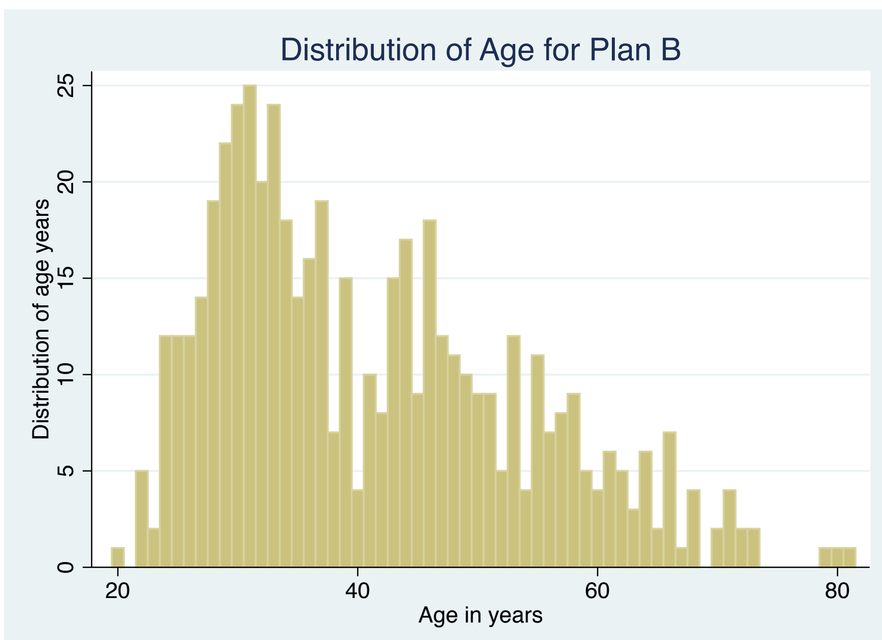

Figure 4: Distribution of age for Plan B

The distribution is more spread out under plan “B”. There is a peak around the age of 35 and another one around the age of 45.

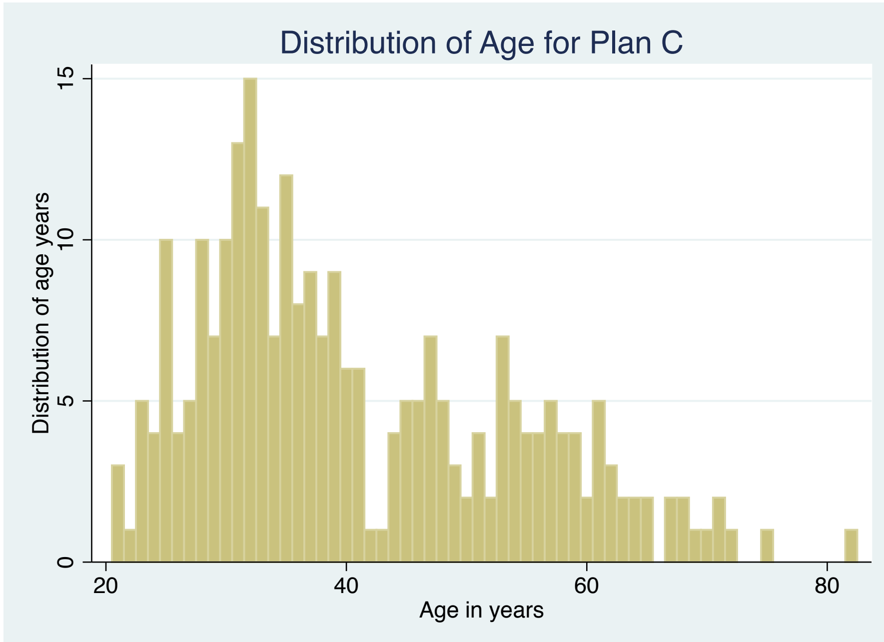

Figure 5: Distribution of age for Plan C

Similarly to plan "B", the distribution of age for members of plan "C" has a peak around the age of 30 and another one around the age of 50.

- Location (*mylocation*)

The location variable is categorical and indicates where a customer picks up the car from. Since some locations had very few rentals compared to others, all observations were grouped in five main areas, namely location V, W, X, Y, Z. The table below shows the number of observations and of customers grouped in each location, along with the percent of late returns for each location. Moreover, the table highlightes the most frequent plan for each location.

Location	Number of observations	Percent of total observations	Percent of late returns within category	Distribution of the plan
V	2,523	26.19%	10.50%	**A (83.95%),** B (9.35%), C (6.70%)
W	1,874	19.45%	7.84%	A (14.99%), **B (49.31%)**, C (35.70%)
X	2,047	21.25%	7.08%	A (23.45%), **B (44.60%),** C (31.95%)

Y	2,392	24.83%	7.40%	**A (61.41%),** B (21.24%), C (17.35%)
Z	797	8.27%	9.03%	A (10.79%), B (40.65%), **C (48.56%)**

Table 2: Location

Table 2 shows that the majority of observations are at location V. This location also has the highest percent of late returns and it is mostly composed of members of plan "A". Because plan "A" has the highest number of late returns among all the plans, it is possible that this factor is affecting the percent of late returns for location V in particular. Plan "B" is mostly found at locations W and X, while plan C at location Z.

- State that issued license (*mylicensestate*)

The variable describing whether a customer received their driving license in state or not is binary and takes on value 0 if the license was issued in-State and 1 otherwise. If a driver is a local, they are more likely to know the area and not get lost. Therefore, switching from being a local to a foreign driver might affect the probability of getting lost and hence returning the car late. The following table shows the number and percent of customers that are either in state or out of state, along with the percent of late returns within each category. Moreover, the table also shows the distribution of the plans within each category.

	Number of customers	Percent of total customers	Number of observations	Percent of total observations	Percent of late return within the plan	Distribution of the plan
In State	929	68.97%	6,457	67.03%	8.32%	A (31.55%), B (37.32%), C (31.13%)
Out of State	418	31.03%	3,176	32.97%	8.47%	**A (75.47%),** B (15.59%), C (8.94%)

Table 3: License State

From Table 3, there is a slightly higher percent of late returns if the driver is out of state, compared to in state. This can be justified by the fact that someone out of state might be less aware of the roads and traffic of the city. Additionally, the distribution of the plans for the in state category is fairly uniform, with plan B being the most frequent at 37.32%. However, plan A is by far the most frequent among out of state customers.

- Membership history (*mymembershiptime*)

This numerical variable indicates how many years a customer has been a member of the company. Because the data contained the day on which a customer became a member, it was straightforward to obtain the number of years since becoming a member. The range of this variable goes from 0 to 9 years. The graph below shows the distribution of the customers based on the years of history with the company.

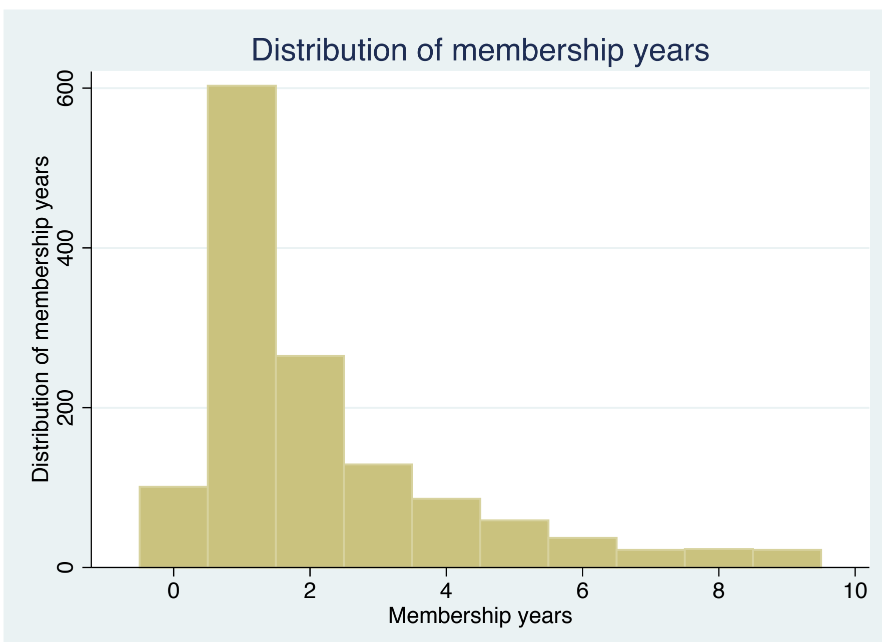

Figure 6: Distribution of membership years

From the graph above, it is noticeable that most of the customers have been members of the company for one year and two years, for a total of 64.44% of the customers. In particular, 603 customers have been members of the company for one year. The longer a customer has been a member of the company, the more they should know about rules of renting and the more experience they should have in selecting an appropriate rental duration. In order to test this, a binary variable was generated. This variable took on value 1 if the return was no more than 15 minutes early and no more than 5 minutes late. Otherwise, the variable was assigned the value 0. It is believed that the longer a customer has been a member of the company and the more they have learned about the system, the more accurate their return time should be. Therefore on one side they should aim at returning no more than 15 minutes early, because if they return more than 15 minutes early, they could have rented the car for 15 minutes less. On the other side, they should not return the car more than 5 minutes late, because doing so would incur a penalty. Then the relationship of this variable to the number of years of the members was studied. The following table shows the percent of accurate returns for each membership year, where an accurate return is a return no more than 15 minutes early or no more than 5 minutes late.

Membership years	0	1	2	3	4	5	6	7	8	9
Percent of accurate returns	71.73%	64.48%	65.79%	64.85%	55.65%	66.92%	65.18%	61.31%	64.49%	74.58%

Table 4: Membership history

Table 4 shows the percent of accurate returns for customers being members of the company from 0 to 9 years. Although there is no clear pattern in the table above, it is noticeable that the higher percent of accurate returns occur when customers have been members of the company for longer.

The third set of variables is related to the choices that customers make every time they make a reservation and attributes related to each experience. They include how long to reserve a car for, the time of day on which a customer makes a reservation, the day of the week the month and whether or not a customer had to pay a penalty fee on their previous rental.

- Reservation time (*myreservationtime*)

Customers can access the booking system online and decide on the duration of their reservation. The variable identifying this is numerical and takes on values ranging from 30 minutes, which is the shortest time a customer can rent a car for, to 105 minutes. The increments of the reservation time are 15 minutes. The table below shows the number and percent of observations in each 15 minutes interval, along with the probability of returning the car late on each category.

Interval	Number of observations	Percent of total observations	Percent of late returns within category	Most frequent type of plan
30 minutes	991	10.29%	10.90%	**A (57.92%),** B (27.04%), C(15.04%)
45 minutes	1,062	11.02%	8.95%	**A (51.60%),** B (30.23%), C(18.17%)
60 minutes	2,381	24.72%	8.44%	**A (44.98%),** B (31.00%), C(24.02%)
75 minutes	1,556	16.15%	7.71%	**A (45.89%),** B (29.76%), C(24.36%)
90 minutes	2,378	24.69%	7.99%	**A (41.80%),** B (32.38%), C(25.82%)
105 minutes	1,265	13.13%	7.27%	**A (42.13%),** B (27.27%), C(30.59%)

Table 5: Rental Time

From the Table above, roughly half of the total number of observations has rental times of 60 or 90 minutes. The highest percent of late returns is for rental of 30 minutes. As the rental time increases, the percent of late returns decreases, from 10.90% for 30 minutes rentals to 7.27% for 105 minutes rentals.

- Time of day (*mytimeofday*)

This variable is categorical and indicates the time of the day that the reservation starts at. After looking at the distribution of rentals throughout the day, the 24-hour period was divided into four groups, each with roughly the same number of observations in them. The table below shows the group number, the time range of the reservation, the number and percentage of observations for each group and the percentage of late returns within each category.

Group	Time of day	Number of observations	Percent of total observations	Percent of late returns within group
1	6am-12pm	2,557	26.54%	6.41%
2	12:15pm-4pm	2,639	27.40%	10.13%
3	4:15pm-7:45pm	2,605	27.04%	7.78%
4	8:00pm-5:45am	1,832	19.02%	8.78%

Table 6: Time of day

Besides for the number of observations, the time of the day upon which the reservation started was taken into account in order to determine the four groups above. Group 1 represents observations of people going to work or to school, group 2 reservations during the day, group 3 people coming out of work or school and group 4 night observations. The highest percent of late returns is related to rentals happening during the day, from 12:15pm to 4pm.

Group	Percent Plan A	Percent Plan B	Percent Plan C	Highest percent of rental duration time
1	35.55%	**36.14%**	28.31%	90 min (26.28%), 60 min (23.07%)

2	**44.30%**	30.24%	25.46%	90 min (27.91%), 60 min (23.92%)
3	**49.72%**	27.31%	22.98%	60 min (26.02%), 90 min (22.57%)
4	**65.64%**	23.08%	11.29%	60 min (27.76%), 30 min (20.32%)

Table 7: Time of day and plan

Table (7) shows that plan "B" is most frequently found during the morning, while during the remaining of the day, rentals are made mostly by members of plan "A".

To test whether this grouping method was good enough, a further analysis was carried out where the cutoff times were shifted by an hour. Therefore group 1 had observations going from 7am to 1pm, group two from 1:15pm to 6pm, and so on. The table below shows the number of observations, percent of total observations and the percent of late returns for this alternative grouping method.

Group	Time of day	Number of observations	Percent of total observations	Percent of late returns within group
5	7am-1pm	3,169	32.90%	6.82%
6	1:15pm-5pm	2,780	28.86%	10.29%
7	5:15pm-8:45pm	2,310	23.98%	7.88%
8	9:00pm-6:45am	1,374	14.26%	8.88%

Table 8: Variation on time of day

Table (8) shows that the number of observations within each group is now not consistent, as it was in the original grouping method. Because the regression run testing this alternative grouping method did not report significant difference with the previous one, the original method was utilized in the final analysis.

- Day of week (*mydayofweek*)

The day of the week for each observation was found using the function *dow* in Stata. This binary variable took on the value 0 if the day of the week was Monday through Friday and 1 if it was Saturday or Sunday. The table below shows the number of observations and percent in each category, along with the duration of the reservation and the type of plan.

Day of week	Number of observations	Percent of observations	Percent of late returns	Plan	Reservation time
M-F (0)	6,656	69.10%	8.35%	**A (45.01%),** B (30.48%), C (24.50%)	**60 (24.20%),** 90 (23.89%)
Sa-Su (1)	2,977	30.90%	8.40%	**A (48.30%),** B (29.43%), C (22.27%)	60 (25.86%), **90 (26.47%)**

Table 9: Day of the week

Table 9 shows that the percent of late returns is roughly the same in spite of a rental happening during the week or during the weekend. Plan "A" is the most frequent in both categories. Moreover, during the week, a higher percent of rentals lasts 60 minutes, while during the weekend it is more common to find rental of 90 minutes.

- Month of the year (*mymonth*)

The months of the year were defined as a categorical variable taking on the values 1 to 12 for the months of January to December respectively.

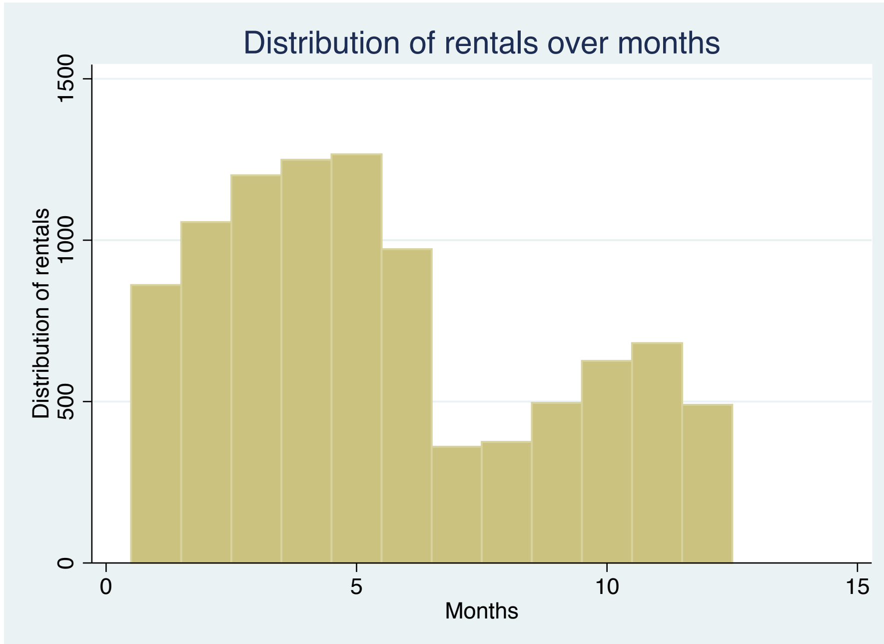

Figure 7: Distribution of rentals over months

The following table shows the pattern for each month of the year. In particular, the type of plan and reservation time is shown for each category.

Month	Type of Plan	Percent late return	Day of the week
January	A (41.11%)	9.99%	M-F (71.78%)
February	A (50.57%)	9.56%	M-F (67.52%)
March	A (48.96%)	7.41%	M-F (64.36%)
April	A (51.40%)	8.25%	M-F (72.38%)
May	A (51.18%)	8.37%	M-F (68.09%)
June	A (39.92%)	8.33%	M-F (69.65%)
July	B (44.17%)	7.22%	M-F (75.00%)
August	B (40.80%)	7.20%	M-F (68.27%)
September	A (41.13%)	6.25%	M-F (66.73%)
October	A (46.96%)	8.15%	M-F (72.52%)

November	A (51.10%)	7.05%	M-F (66.67%)
December	A (47.55%)	11.63%	M-F (70.20%)

Table 11: Month

Table 11 shows that plan "A" is the most commonly found throughout the months, with the exception of July and August. The percent of late returns is higher for the months of December, January and February. This factor might be due to more harsh weather conditions over the winter season.

- Penalty (*mypenalty*)

This binary variable indicates whether a customer had to pay a penalty on their previous reservation. The variable takes value 1 if a penalty occurred and 0 otherwise. This value was assigned by sorting the observations first by member id and then by reservation day. If the return on a reservation was late, then the penalty on the following reservation was 1; otherwise it was 0. This variable is believed to affect the decision upon the reservation time. In fact, if a customer incurred a penalty fee on their previous car rental due to a late return, they will be more likely to reserve the car for longer on their next reservation, so not to incur a fee again. The following table shows the number and percent of rentals that incurred the fee.

Penalty	Number of observations	Percent of observations
Yes	747	7.75%
No	8,886	92.25%

Table 12: Penalty

Only 7.75% rentals of the total number of observations had to pay a penalty. The 747 rentals were made by a total of 350 customers. The customer that paid the most number of penalties incurred 24 penalties out of their 127 transactions.

The fourth group of variables has to do with those attributes that are not controllable either by the company or the customers. Such variables are the weather conditions and temperature on the rental days. In order to get the desired information, a file containing the weather history of the desired city was downloaded from online. The file was found simply searching online. The file contained daily, weekly, and monthly data of average temperature, wind chill, humidity, pressure, visibility, wind speed events and condition. The file was merged according to the date.

- Weather conditions (*myweatherconditions*)

From the merged file, a binary variable was generated. The file contained the condition of the weather on each day and recorded conditions of “Snow”, “Rain” and “Fog”. This variable took on value 1 if the weather was snowy, rainy or foggy and 0 if it was sunny. The effects of snow, rain and fog are believed to have the same consequences on customers’ perceptions of the weather and were therefore all coded together. The table below reports the percent of days with snow, rain and fog for each month.

Month	Percent of days with snow, rain or fog	Percent of late returns
January	59.67%	9.99%
February	44.64%	9.56%
March	37.10%	7.41%
April	63.33%	8.25%
May	50%	8.37%
June	60%	8.33%
July	32.26%	7.22%
August	19.35%	7.20%
September	40%	6.25%
October	54.84%	8.15%
November	23.33%	7.05%

December	74.19%	11.63%

Table 13: Weather and month

December was the month with the highest percent of snow, rain or fog and also the highest percent of late returns. However, other months do not seem to have a clear relationship between the percent of bad weather and late returns.

- Weather temperature (*mytemperature*)

This numerical variable indicates how many degrees were there on each rental day. The range of values went from -17F to 89F. The temperature given in the downloaded file was an average for that day.

Lastly, variables indicating a return on time and a return more than 15 minutes late will be here defined once again.

- Return on time (*y*)

This binary variable indicates whether a return was on time or late. If a customer returned a car on time, the variable took on value 0. If the return was late, the value assigned to the variable was 1. If a customer returned the car within five minutes of their reservation end time, the reservation was considered on time because the customer did not incur any penalty fee. Therefore, in the eyes of both the customers and the company, returning within five minutes is not considered a late return. The following table shows the number of observations and percent of on-time and late returns.

	Number of observations	Percent of observations
Return on time	8,827	91.63%
Return late	806	8.37%

Table 14: Return on time and late

Table 14 shows that most rentals are returned on time, with only 8.37% of observations being late.

- Return more than fifteen minutes late (z)

This is a binary variable indicating whether a customer returned the car very late or not. If the return was more than fifteen minutes late, the variable took on value 1; otherwise it had value 0. Fifteen minutes is considered to be the cutoff time because if a customer returns the car later than that, then the company will have to take action in order to assure that the following rentals will not be negatively affected. The table below shows the percent of returns within 15 minutes and over 15 minutes.

	Number of observations	Percent of total observations
Return within 15 minutes	9,387	97.45%
Return after 15 minutes	246	2.55%

Table 15: Return more than 15 minutes late

The percent of very late return is very small, only 2.55%. This means that the majority of customers return the car within fifteen minutes of their reservation end time.

6 Results and Discussion

After defining the explanatory variables and coding them, the random effect logistic regression was run with Stata. This section reports the results of the regression, along with a discussion of which variables affect late returns the most. A fixed effect regression was run for both y and z and the outcome is reported below.

```
Conditional fixed-effects logistic regression   Number of obs      =      4990
Group variable: memberid                        Number of groups   =       324

                                                Obs per group: min =         2
                                                               avg =      15.4
                                                               max =       127

                                                LR chi2(26)        =     68.08
Log likelihood  = -1473.7326                    Prob > chi2        =    0.0000
```

Figure 8: Fixed effects for *y*

From Figure 8, the p-value of the Likelihood Ratio (LR) Chi-Square is 0.

```
Conditional fixed-effects logistic regression   Number of obs      =      2839
Group variable: memberid                        Number of groups   =       145

                                                Obs per group: min =         2
                                                               avg =      19.6
                                                               max =       127

                                                LR chi2(26)        =     37.23
Log likelihood  = -525.71861                    Prob > chi2        =    0.0712
```

Figure 9: Fixed effects for *z*

Running the fixed effects regression for *z* yields the much higher p-value of 0.0712.

The random effects regression was then run for both *y* and *z*. The figures below show the outcomes of this regression.

```
Random-effects logistic regression              Number of obs      =      9633
Group variable: memberid                        Number of groups   =      1303

Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       7.4
                                                               max =       127

Integration method: mvaghermite                 Integration points =        12

                                                Wald chi2(28)      =     96.41
Log likelihood  = -2528.8767                    Prob > chi2        =    0.0000
```

Figure 10: Random effects for *y*

```
Random-effects logistic regression              Number of obs      =      9633
Group variable: memberid                        Number of groups   =      1303

Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       7.4
                                                               max =       127

Integration method: mvaghermite                 Integration points =        12

                                                Wald chi2(28)      =     51.56
Log likelihood  = -1062.5452                    Prob > chi2        =    0.0043
```

Figure 11: Random effects for *z*

The p-values of the Wald Chi-Square test for *y* and *z* are 0 and 0.0043 respectively. As stated earlier, a random effects logistic regression was chosen because it better describes the data and output of this project.

In order to understand which variables affect late returns the most, odd ratios are provided. Odd ratios indicate the association between two outcomes; in particular how changing one variable by one unit affects the percent change of the dependent variable. The table below shows the odd ratios for both late returns and returns more than fifteen minutes late for all of the explanatory variables.

Variables	**Variation in Probability of returning on time**	**Variation of Probability of returning more than 15 minutes late**
Type of plan	-1.12%	0.30%
Per unit rental time	-0.06%	0.65%
Age	-1.28%	-1.44%
Location (*V*)		
W	-5.35%	-25.25%
X	-13.71%	-5.99%
Y	-27.64%	-14.54%
Z	-7.14%	-11.57%
State that issued license	-11.77%	-25.13%
Membership history	-7.27%	-6.73%
Reservation time	-0.32%	-0.63%*
Time of day (*6am-12pm*)		
12:15pm-4pm	**68.77%** *(62.15%)*	**60.45%*** *(66.17%)*
4:15pm-7:45pm	9.80% *(6.54%)*	9.77% *(12.34%)*

8:00pm-5:45am	1.43% *(0.90%)*	22.85% *(17.85%)*
Day of week	-1.20%	10.80%
Month (*January*)		
February	0.27%	2.51%
March	-32.41%*	-0.77%
April	-26.87%	8.88%
May	-26.78%	34.45%
June	-21.98%	45.68%
July	-37.09%	**64.04%**
August	-40.42%	17.94%
September	-47.20%	4.59%
October	-23.05%	42.16%
November	-41.62%*	-23.08%
December	17.21%	32.72%
Penalty	**55.49%***	**90.55%***
Weather conditions	4.29%	**26.77%**
Weather temperature	0.24%	-0.92%

Table 16: Odd Ratios. The * means significance at the 5% level. The bold means a high percent change. The number in parenthesis mean the odd ratios when the alternative grouping method was utilized.

The table above shows the change in the probability of returning the car late, along with the significance of each variable at the 5% level. The odds ratios show the change in percent from the base level, which is the lowest number. A discussion will be carried out first of the variables affecting the probability of late returns and then of the variables affecting the probability of returning more than fifteen minutes late.

Most variables have minor percent change on whether a customer will return a car on time or late. In particular, varying variables such as type of plan, per unit rental time, age, reservation time, day of the week, weather conditions and temperature causes a less than five percent change respectively on the outcome of returning the car late. Variables having a bigger impact are the location, the time of the day, the month and the penalty. Additionally, the time of the day and the penalty are statistically significant and allow for the rejection of the null hypothesis, which was that customers are rational when making rental decisions. When customers switch from renting a car in the morning to the afternoon time, the percent change of a late return is 68.77%. This means that they are 68.77% more likely to return the car late. This factor might be due to the association in customers' mind of higher demand during the day. Similarly, if customers paid a penalty on their previous rental, they are 55.49% more likely to return the car late on their following rental. This change might be due to the behavior of the customers and the negative effect of paying a penalty fee. In fact, this factor is significant of irrationality and therefore customers might not learn from previous experience.

When analyzing returns more than fifteen minutes late, it is still noticeable that type of plan, per unit rental time, age and temperature slightly affect the percent change on late returns. The location negatively impacts the percent of late returns, meaning that changing reservation from one location to another doesn't positively increase the percent of late returns. Again, time of day and penalty are statistically significant to whether a customer returns a car more than fifteen minutes late or not. Weather conditions change the probability of a very late return by 26.77%. This means that when customers return a car late due to bad weather conditions, they are more likely to return it more than fifteen minutes late. The month variable also affects the probability of late returns. In particular, renting a car in July as opposed to January increases the probability of returning more than fifteen minutes late by 64.04%.

7 Conclusion and Future Work

Overall, the random effect logistics analysis showed that the time of day and the penalty are the two most critical variables affecting the probability of a customer to return the car late. From this research, it is possible to identify a specific category of customers, namely those who rent a car during the day, from 12:15pm to 4pm, and had to pay a penalty fee on their previous rental, to be more likely to return the car late. This insight suggests that the history of the members affect how customers see their future rentals and therefore should be taken into account by the company when designing contracts for customers. Other variables have a minimal impact on the percent of late returns. Future work includes coming up with an optimal contract design for customers reserving in the range of day where most late returns happen, or for customers who paid a penalty fee on their previous rentals. Moreover, it would be interesting to test the interaction of some of the variables and how those interactions affect the percent of late returns. For the purposes of this research, interactions were not included in the model. However, combining type of plan and day of the week might show some patterns of how customers make rental decisions and whether they are rational in doing so.

References

S. Aflaki and I.Popescu, "Managing Retention in Service Relationships Managing Retention in Service Relationships," 2011.

F. Bardhi and G. M. Eckhardt, "Access-Based Consumption: The Case of Car Sharing," *J. Consum. Res.*, vol. 39, no. 4, pp. 881–898, Dec. 2012.

W. J. Carroll, R. C. Grimes, "Evolutionary Change in Product Management: Experiences in the Car Rental Industry"**,** *Interfaces*, Vol. 25, No. 5, pp. 84-104, Sep. - Oct. 1995.

S. S. Iyengar, M. R. Lepper, "When choice is demotivating: Can one desire too much of a good thing?" *J. Personality Soc. Psych*. 29(6) 995–1006, 2000.

P. C. Reiss, F. A. Wolak, "*Structural Econometric Modeling: Rationales and Examples from Industiral Organization*", Handbook of Econometrics, Vol. 6A, 2007.